

SVEUČILIŠTE U ZAGREBU  
FILOZOFSKI FAKULTET  
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI  
Ak. god. 2015./ 2016.

Laura Bezić

**PREDNOSTI I MANE PRIKUPLJANJA VELIKIH PODATAKA**

Završni rad

Mentor: dr.sc. Kristina Kocijan, doc.

Zagreb, 2016.

## Sadržaj

1.	Uvod.....	2
2.	Veliki podatci.....	3
3.	Prikupljanje Velikih podataka .....	5
4.	Datafikacija.....	6
5.	Privatnost .....	9
6.	Prednosti i mane Velikih podataka .....	13
6.1.	Prednosti.....	18
6.1.1	Poboljšanje proizvodnje i proizvoda.....	18
6.1.2.	Predviđanje kriminala .....	20
6.1.3	Medicinski napredak.....	21
6.1.4.	Obrazovanje .....	23
6.2.	Mane.....	24
6.2.1.	Gubitak privatnosti i nadzor .....	25
6.2.2.	Loš rezultat predviđanja .....	28
6.2.3.	Stvaranje društvenog jaza.....	30
7.	Zaključak .....	32
8.	Literatura.....	34

## 1. Uvod

U ovom radu bit će riječ o Velikim podacima, točnije o prednostima i manama njihovog prikupljanja. Iako su svoj uspon Veliki podatci doživjeli 2008. godine, van stručne zajednice o njima se ne govori puno (Kocijan, 2014). Kroz studij Informacijskih i komunikacijskih znanosti došla sam u dodir s Velikim podacima, a time i shvatila njihovu vrijednost i važnost. Kao poticaj za odabir upravo ove teme za završni rad bila su mi začuđena lica ljudi kada bi se spomenuli Veliki podatci. Čini mi se da je čak i onima upućenima u Informacijske znanosti i područje računarstva promakla ova revolucija. Stoga ovim radom želim pokušati približiti što većem broju ljudi koliko su Veliki podatci dio naše svakodnevice i kako smo skoro svi njihovi korisnici.

U prvom ću poglavlju obrazložiti pojam Velikih podataka i njegove karakteristike. Potom ću opisati kako se prikupljaju Veliki podatci i na koji se način prikupljanje podataka promijenilo u odnosu na ranije te ću se osvrnuti na značenje datafikacije. U poglavlju o privatnosti će biti riječi o povezanosti privatnosti i Velikih podataka, zakonima koji se vežu uz ovaj problem u Hrvatskoj i svijetu, te o neinformiranosti korisnika o njihovoj privatnosti. Peto poglavlje kao glavno obuhvaća mane i prednosti prikupljanja Velikih podataka dajući objašnjenje što će se u radu smatrati manama i prednostima prikupljanja Velikih podataka te njihovo slikovito pojašnjenje kroz dva grafička prikaza. Kroz primjere prikupljanja Velikih podataka pokušat ću dodatno pojasniti mane i prednosti prikupljanja. Rad ću završiti kratkim zaključkom, te osobnim mišljenjem u kojem smjeru bi priča o Velikim podacima mogla ići u budućnosti.

## 2. Veliki podatci

Prije same definicije pojma Velikih podataka pozvala bih se u ovom poglavlju na autoricu Kristinu Kocijan (2014). Slijedeći njezine argumente o brojnim mogućnostima koje ovaj pojam otvara odlučujem se da i sama prihvatim pisanje pojma velikim početnim slovom.

U svojoj biti Veliki podatci govore o predviđanju. Iako su opisani kao dio računalnih znanosti, kao dio umjetne inteligencije, točnije unutar područja strojnog učenja, ovakva karakterizacija je varljiva. Kod Velikih podataka nije bit u pokušaju da se uči računalo da „misli“ kao čovjek, već se radi o primjeni matematike na velike količine podataka kako bi se moglo doći do zaključka o određenoj mogućnosti, na primjer koja je vjerojatnost da je određena elektronička pošta nepoželjna ili da su otipkana slova *teh* trebala biti *the* (Mayer-Schönberger, Cukier; 2013).

Daljnje objašnjenje pojma Velikih podataka često se povezuje s 3 V: volumen, velicitet (brzina prikupljanja) i varijantnost (raznolikost tipova podataka). Pri tome govorimo da su *Veliki podatci enormni u volumenu jer sadržavaju terabajte ili petabajte podataka, velikog su veliciteta tj. brzine jer su stvarani u skoro stvarnom vremenu te raznoliki u varijantnosti tipova jer su strukturirani i nestrukturirani u prirodi i često privremeno i prostrano povezivani* (Kitchin, 2014: 68)<sup>1</sup>.

Pojam Velikih podataka bio je korišten u znanosti te se referirao na velike skupove podataka koji su zahtijevali superračunala. No, ono za što su nekad bila potrebna superračunala, danas se može analizirati i na osobnim računalima sa standardnim softverom. Nema sumnje da je kvantiteta podataka koji su trenutno dostupni poprilično velika, ali to nije temelj definiranja novog ekosistema podataka. U biti, neki podatci obuhvaćeni Velikim podacima uopće nisu toliko veliki, koliko su bili veliki raniji

---

<sup>1</sup> U originalu: Big data are:

- huge in *volumen*, consisting of terabytes or petabytes of data;
- high in *velocity*, being created in or near real-time;
- diverse in *variety* in type, being structured and unstructured in nature, and often temporally and spatially referenced

skupovi podataka koji se ne smatraju Velikim podacima. Kod Velikih podataka od njihovog obujma važniji je njihov kapacitet za traženje, grupiranje i unakrsne reference velikih skupova podataka (Boyd, Crawford; 2012).

Prema Boydu i Crawfordu (2012) Veliki se podatci definiraju kao kulturni, tehnički i školski fenomen koji počiva na međusobnom djelovanju:

1. **tehnologija:** maksimizacija snage izračunavanja i algoritmičke točnosti skupljanja, analize, vezivanja i uspoređivanja velikih skupova podataka;
2. **analiza:** nacrt na temelju velikih skupova podataka kako bi se identificirali obrasci iz kojih se rade ekonomski, socijalni, tehnički i legalni zahtjevi;
3. **mitologija:** široko rasprostranjeno vjerovanje kako veliki skupovi podataka nude višu formu inteligencije i znanja koji mogu generirati uvide koji su prije bili nemogući, sa aurom istine, objektivnosti i točnosti.

### 3. Prikupljanje Velikih podataka

Prikupljanje Velikih podataka u današnje doba otvara tvrtkama, državnim ustanovama i ostalim organizacijama mogućnosti za uspjeh kod korisnika. Kroz individualan pristup stvara se kvalitetnija usluga čime je korisnik zadovoljniji, nego u slučaju kada mu se pristupa globalno. Pogledamo li kroz povijest, primijetiti ćemo da su se podatci prikupljali još u drevnim egipatskim i kineskim civilizacijama. U 11. stoljeću velika količina podataka prikuplja se u *Knjizi sudnjeg dana*, britanskom imovinskom popisu, dok u najnovijoj povijesti poznajemo sustavne demografske popise. Pri tome, ovakvo prikupljanje podataka odlikuju dugotrajni procesi, djelomični podatci te u konačnici neaktualni rezultati anketa (Kocijan, 2014).

U digitalnoj eri pri prikupljanju Velikih podataka spremni smo žrtvovati potpunu točnost kako bismo saznali aktualan trend. Tako će se nekad u ovom procesu dogoditi da dva i dva daju 3.9 i na to pristajemo.<sup>2</sup> Bitno je shvatiti da Velikim podacima nije svojstven nered, već je on prisutan kod podataka koje koristimo pri mjerenjima, snimanjima i analizi informacija. Pri tome se nered pojavljuje u raznim oblicima: kod spajanja više izvora informacija dolazi do nesavršenog usklađivanja, ponekad formati podataka moraju biti prepravljani prije obrađivanja, a kod samog obrađivanja podatci se transformiraju i pretvaraju u nešto drugo, čime se ponovno može stvoriti nered u formatima. Kada bi tehnologija odjednom postala savršena, nepreciznost bi nestala. Dok se to ne dogodi, nered je nešto s čime se moramo nositi. Inzistirati na točnosti nema ekonomskog smisla jer je neodoljivije imati što veću količinu podataka (Mayer-Schönberger, Cukier, 2013). Isto tako Veliki su podatci dostupni svima koji za njih pokažu interes, a ne samo pripadnicima društvenih znanosti, čime se mijenja i sam proces prikupljanja podataka (Boyd, Crawford; 2012).

---

<sup>2</sup> Autor ovim slikovito želi reći kako nam Veliki podatci ponekad daju tek približni izračun onoga što želimo znati, ali taj izračun bi sam po sebi trebao biti dovoljan.

Ovim načinom prikupljanja podatka društvo je pristalo na dva kompromisa. Zbog velike količine podatka koji se prikupljaju, smatra se da se ne mogu koristiti u njihovoj potpunoj količini, te se zadovoljava približnim  $N=svi$ . Kao drugi kompromis pojavljuje se kvaliteta informacija. Privilegija točnosti pripada malim podacima, prije se pri prikupljanju manje količine informacija, očekivalo da njihova točnost bude najveća moguća. Danas se od nas očekuje da prihvatimo zamagljenu sliku rezultata istraživanja kako bismo na kraju vidjeli širu sliku. Zahtjeva se i prilagođavanje na nered i nesigurnost jer kada prihvatimo da ne postoji samo jedan odgovor na zadano pitanje koje tražimo, bit ćemo bliže realnosti (Mayer-Schönberger, Cukier; 2013).

### 4. Datafikacija

S datafikacijom, informacija opet postaje primarna u informacijskoj tehnologiji. Prikupljaju se informacije o svemu što nas okružuje te se one dalje transformiraju kako bi se mogle prebrojati i analizirati, čime se otvaraju brojne nove mogućnosti (Kocijan, 2014). *Dataficirati znači staviti u brojivi format kako bi se moglo tabelirati i analizirati* (Mayer-Schönberger, Cukier; 2013:78)<sup>3</sup>. Kako bismo mogli provesti datafikaciju informacija, moramo znati kako mjeriti i kako snimiti ono što mjerimo te imati pravi alat.

Kako bi se mogla dobiti jasnija slika pojma datafikacije, potrebno je poslužiti se primjerima koji pokazuju na koji način prikupljanje i analiziranje informacija može biti od koristi. Jedan od primjera primjene datafikacije razvoj je novog sustava zaštite od krađe automobila. Japanski profesor Shigeomi Koshimuzu uzeo je podatke vozačevog sjedenja u automobilu i proizveo sustav suzbijanja krađe zasnovan na kutu sjedenja u automobilu koji je karakterističan za svakog vozača. Sustav pri tome utvrđuje točnost digitalnog koda na temelju kuta sjedenja. Tako je primaran podatak o poziciji tijela iskorišten u svrhu sprječavanja krađe (Mayer-Schönberger, Cukier; 2013).

Originalnu primjenu datafikacije pokazao je Google u svome projektu Google Books. Iako Google Books kao projekt nije unikatna te je većini poznat i projekt Gutenberg koji omogućuje čitateljima besplatno čitanje knjiga kao i drugi slični portali,

---

<sup>3</sup> U originalu: To datafy a phenomenon is to put it a quantified format so it can be tabulated and analyzed.

Google je jedini promatrao štivo kao datafikaciju. Sadržaj knjiga iskoristio je ne samo kao materijal za čitanje i analiziranje popularnosti, već i za usavršavanje Google prevoditelja. Kako su knjige često prevedene na više jezika, Google je prijevode tretirao kao veliki matematički zadatak, pri čemu je računalo trebalo odlučiti vjerojatnost riječi koja daje najbolji prijevod (Mayer-Schönberger, Cukier; 2013). Koliki je odaziv izazvao projekt Google Books pokazuje sudski proces koji je zbog ovog projekta pokrenut protiv tvrtke Google 2005. godine, a čija je presuda donesena u travnju 2016. U sudskom je procesu društvo autora *Authors Guild* tužilo tvrtku Google zbog povrede autorskih prava. Nakon više od deset godina vrhovni sud Sjedinjenih Američkih Država odlučuje kako su skenirani dokumenti koje Google koristi legalni i da Google Books ostaje u upotrebi. Odluka je objašnjena tezom da knjige služe za javni interes i da njihovom upotrebom u sustavu Google Books nisu ugrožene. Isto tako je odlučeno kako knjige nisu podložene transformaciji jer pri ovom procesu njihovom sadržaju se ne dodaje ništa novo niti se komentira sadržaj digitaliziranih tekstova (Kuhn, 2016).

Twitter je također jedan od primjera datafikacije koji je koristi i za svoju dobit, prodajom podataka o *tweet*-ovima. Jedna od primjena *tweet*-ova je analiza uspjeha filma prije izdavanja filma a na temelju reakcija korisnika Twittera. Bez obzira što *tweet*-ovi ne sadrže puno informacija već samo 140 znakova, softver koristi meta podatke, podatke o podacima, čime omogućuje puno širu korist onoga što prikupi (Mayer-Schönberger, Cukier; 2013). Za razumijevanje ovog primjera možemo uzeti najobičniji primjer traženja datoteke na računalu. Zahvaljujući meta podacima, računalo unutar jednog dokumenta pamti lokaciju tog dokumenta, autora, naslov, ključne riječi, opis, datum stvaranja i uređivanja te slične korisne informacije koje nam olakšavaju traženje. Isto je tako, na temelju *tweet*-ova, moguće uzeti informaciju primjerice koja je dobna skupina osobe koja je objavila *tweet*, gdje ta osoba živi, jesu li riječi koje koristi negativne ili pozitivne i sl. Ti podatci se potom mogu iskoristiti na različite načine, a ovisno o varijablama pretrage.

Datafikacija ne prikazuje samo pojedinačne stavove ili osjećaje u obliku koji se može analizirati, već i načine ponašanja pojedinaca ili društvenih skupina. Različite društvene mreže i Google posjeduju blago podataka čijom datafikacijom, a kasnije



analizom, možemo doći do zaključaka o različitim društvenim obrascima ponašanja pojedinca ili društva u cjelini (Mayer-Schönberger, Cukier, 2013).

Datafikacija će zasigurno promijeniti način na koji promatramo svijet oko sebe. Putem Interneta ona već utječe na kreiranje naših svakodnevnica. Datafikacija shvaća svijet kao oceane informacija te omogućuje da istražimo njihova prostranstva i dubine (Mayer-Schönberger, Cukier, 2013).

## 5. Privatnost

Privatnost i Veliki podatci percipiraju se često kao suprotnosti. Važnost Velikih podataka i njihov potencijal leži u njihovom prikupljanju i sve moćnijim načinima da se ti podatci analiziraju, a privatnost vežemo uz povjerljivost, uz područja života koja želimo zadržati za sebe i posjedovati kontrolu nad njima. Kada govorimo o privatnosti, bitno je naglasiti da Internet pri svom nastajanju nije zamišljen na način kakvim se danas upotrebljava. Internet je zamišljen kao mreža za dijeljenje informacija, a ne za skrivanje podataka od javnosti te zbog toga nisu postavljeni temelji za sigurnost i privatnost (Reno, 2012).

Definicija privatnosti s pojavom Interneta bitno se mijenja. Osobni podaci prikupljaju se i obrađuju na različite načine, a kamen je temeljac koncept obavijesti i suglasnosti (eng. *notice and consent*<sup>4</sup>) koji je zakonit i poštuje princip privatnosti. Međutim, s Velikim podacima, on gubi na značenju. Ne možemo očekivati da se korisnici unaprijed usuglase s nečim što još ne postoji. Analitičari tvrde da bi se o svakom ponovnom korištenju privatnih podataka trebalo obavijestiti pojedinog korisnika te od njega tražiti pristanak za korištenje informacija. No, niti jedna kompanija ne bi podnijela troškove takvog pothvata. Druga alternativna je razmišljanje unaprijed. U tom bi se slučaju od korisnika tražilo da pristanu na sve možebitne buduće upotrebe njihovih podataka. Međutim, to bi uništilo pravo korisnika na privatnost koju ostvaruje davanjem suglasnosti (Mayer-Schönberger, Cukier, 2013).

Europska je unija 2015. godine donijela zakon o zaštiti osobnih podataka <sup>5</sup> na temelju kojega pravne osobe poput tvrtki Google, Amazon i Facebook moraju jasno zatražiti pristanak korisnika kako bi koristili njihove podatke. Ovim zakonom uvodi se i pravo korisnika na zahtjev za brisanje osobnih podataka i premještanje podataka od jednog

---

<sup>4</sup> Koncept obavijesti i suglasnosti je formaliziran 1998, a pod njim su se podrazumijevala pet osnovnih pravila zaštite privatnosti: obavijest, dostupnost, suglasnost, naknada, integritet i sigurnost. WPressUTexas (2015)

<sup>5</sup> Originalan naziv zakona: Directive 95/46/EC (General Data Protection Regulation), dostupan na: [http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:ST\\_5419\\_2016\\_INIT](http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:ST_5419_2016_INIT)

pružatelja usluga drugome. Nadalje se zahtjeva od tvrtki da prilagode svoje usluge i prikupljaju što manje podataka korisnika. Isto tako se omogućuje korisniku da sazna koji se podatci o njemu prikupljaju i u koje svrhe. Zakon se odnosi i na tvrtke iz Sjedinjenih Američkih Država koje se moraju pridržavati ovog zakona, ako svoje usluge žele nuditi u Europskoj uniji. Propisane kazne za kršenje zakona izriču se u milijardskim iznosima, točnije 4% godišnje dobiti tvrtke koja prekrši zakon. Ovaj bi zakon trebao stupiti na snagu 2018. godine kada bi ga trebale usvojiti sve članice Europske unije čime bi se riješio problem prijašnjeg Zakona o zaštiti podataka u Europskoj uniji iz 1995. godine, kada je svaka članica imala pravo prilagodbe propisa, a što je omogućilo tvrtkama da odaberu sjedište u onoj članici koja ima najmanje kazne za kršenje zakona. Novi zakon ima jednu iznimku, a to su Veliki podatci, čije bi se prikupljanje i upotreba dozvolili samo za znanstvene svrhe. Ova nadopuna bi trebala biti uvrštena u zakon u prvoj polovici 2018. godine (Ackermann, 2015).

Kada govorimo o hrvatskim odredbama, za sada imamo Zakon o zaštiti osobnih podataka i Zakon o elektroničkim medijima. Zakonom o zaštiti osobnih podataka se definiraju osobni podatci kao *svaka informacija koja se odnosi na identificiranu fizičku osobu ili fizičku osobu koja se može identificirati* (Narodne Novine, 2012:Članak 2.). Zakon se osvrće na obradu podataka, pri čemu se obrađivanje podataka dozvoljava isključivo prema odredbama ovog zakona, od kojih glavna odredba glasi: *Osobni podaci mogu se prikupljati u svrhu s kojom je ispitanik upoznat, koja je izričito navedena i u skladu sa zakonom i mogu se dalje obrađivati samo u svrhu u koju su prikupljeni, odnosno u svrhu koja je podudarna sa svrhom prikupljanja. Daljnja obrada osobnih podataka u povijesne, statističke ili znanstvene svrhe neće se smatrati nepodudarnom, pod uvjetom da se poduzmu odgovarajuće zaštitne mjere.* (Narodne novine, 2012: Članak 6.) U daljnjem je tekstu pojašnjeno kako podatci moraju biti bitni za svrhu prikupljanja, opseg prikupljenih podataka ne smije biti većeg opsega nego je to nužno, podatci moraju biti točni, potpuni i ažurni, te se očekuje odgovarajuće čuvanje podataka. Zakonom je jasno propisano da je izvršitelj obrade prikupljanja podataka prije njihovog prikupljanja, dužan informirati osobu čiji se podatci prikupljaju. U daljnjem tekstu je nejasno je li u slučaju prikupljanja podataka u znanstvene ili statističke svrhe osobu potrebno

informirati. Zakon ostavlja mogućnost da se osoba koja smatra da su njezini podatci nepravomoćno prikupljeni žali Agenciji za zaštitu osobnih podataka. Za povredu zakona predviđene su kazne od 20 000 do 40 000 kuna (Narodne novine, 2012). Iako se u zakonu o elektroničkim medijima pojavljuju internetske stranice, Internet kao medij nije obuhvaćen samostalno, niti postoji osvrt na zaštitu podataka unutar njega. Iako u Ustavu republike Hrvatske pravo na privatnost nije jasno definirano već samo navedeno: *Svakom se jamči štovanje i pravna zaštita njegova osobnog i obiteljskog života, dostojanstva, ugleda i časti* (Narodne novine, 2004: Članak 35.). Ovo pravo nalazimo u Zakonu o medijima kao objašnjenje pojma privatnosti, pa se može tumačiti i kao definicija tog pojma.

Dok će se Hrvatska u narednim godinama prilagoditi europskim zakonskim okvirima, SAD ima drugačiju sliku privatnosti. Europska slika privatnosti vezana je uz državu koja brani interese građana zakonima, dok se u Sjedinjenim Američkim Državama veća vrijednost polaže na slobodu govora. Svaka država za sebe donosi zakon koji bi trebao osigurati privatnost, a oni se često značajno razlikuju, što rezultira time da tvrtke posluju prema zakonu države u kojoj imaju sjedište, a sjedište mogu birati prema zakonu koji im više odgovara. U državama poput Kine ili Vijetnama, privatnost nestaje pod strogom nadležnosti političkog režima, dok se u Japanu ili Južnoj Koreji ona strogo čuva zakonom (Laboratorij za sustave i signale, 2010).

Jedan od primjera povrede privatnosti veže se uz Facebook kao jednu od tvrtki koja raspolaže zavidnom količinom Velikih podataka. Tako je početkom 2014. protiv tvrtke Facebook podignuta tužba pod sumnjom da ova tvrtka povrjeđuje pravo na privatnost time što čita poruke korisnika bez njihovog pristanka. Sumnjiče se i da su tako prikupljene podatke prodali trećoj strani (Hofstetter, 2014).

Sličan su primjer i banke koje nastoje prikupiti što više podataka o svojim korisnicima. Međutim, te prikupljene podatke banke uglavnom ne aktualiziraju pa se o korisniku stvara slika koja nije nužno istinita. Nameće se logično pitanje tko ima pravo i pristup tim podacima? Iako je jasno da je prodaja podataka unosan posao, za sada nitko nije dao odgovor tko je odgovoran za to. Autorica Yvonne Hofstetter (2014) definira fenomen digitalnog zombija kao našeg klona koji postoji u virtualnom svijetu.

Karakteriziraju ga podatci koji su o nama prodani te se koriste za analizu, no oni ne moraju biti nužno točni. Autorica vidi problem u tome što korisnici nisu svjesni karakteristike svog digitalnog zombija.

Upravo ta nesvjesnost dovodi do problema da je ljude jako teško zaštititi od zloupotrebe Velikih podataka. Većina korisnika ne vidi nikakvu prijetnju i nisu spremni poduzeti radnje kako bi zaštitili sami sebe, posebice kada se radi o odricanju od besplatnih usluga poput elektroničke pošte, internetskih tražilica ili društvenih mreža. Kako većina preglednika nema jednostavnu zaštitu od praćenja korisnika, Europska je unija još 2009 uvela direktivu o kolačićima (eng. *cookies*) koja nalaže da tvrtke na internetskim stranicama jasno naglase da koriste kolačiće, da upozore korisnika i da traže njegov pristanak za prikupljanje njegovih podataka (EU direktiva o kolačićima, 2009). Nadalje bitna je sastavnica svake internetske stranice dokument o zaštiti privatnosti, ali on često ne poboljšava situaciju. Većinom je riječ o pravnim dokumentima koji svojom duljinom, stilom, stručnim terminima i složenosti, potiču korisnike da ih jednostavno preskoče. Reno (2012) smatra kako pri tome korisnici razmišljaju da će vlasnici stranica naposljetku ipak iskoristiti njihove podatke zaštićeni pravnim dokumentima. Vlasnicima internetskih stranica je to i omogućeno kroz pravo mijenjanja objašnjenja o zaštiti podataka. Većina društvenih mreža ima postavke privatnosti koje korisnici mogu podesiti. Što predstavlja korak naprijed, ali postavke su često originalno podešene tako da se ne zaštićuje korisnik. Prema istraživanjima, malo korisnika društvenih mreža samostalno mijenja ove postavke. Tako prema istraživanju Pew Research centra iz 2011., 48% korisnika kaže kako imaju probleme s namještanjem sigurnosnih postavki, dok druga polovica drži profil otvoren samo za prijatelje.<sup>6</sup> Također, sami načini na koji se podatci prikupljaju često su tehničke prirode. Tako većina ljudi ne zna što su kolačići i da se oni koriste za prikupljanje podataka. Ako uzmemo u obzir da postoje načini kojima bi se internetske korisnike o tome moglo informirati, očigledno je da za to, ustvari, nema interesa (Reno, 2012).

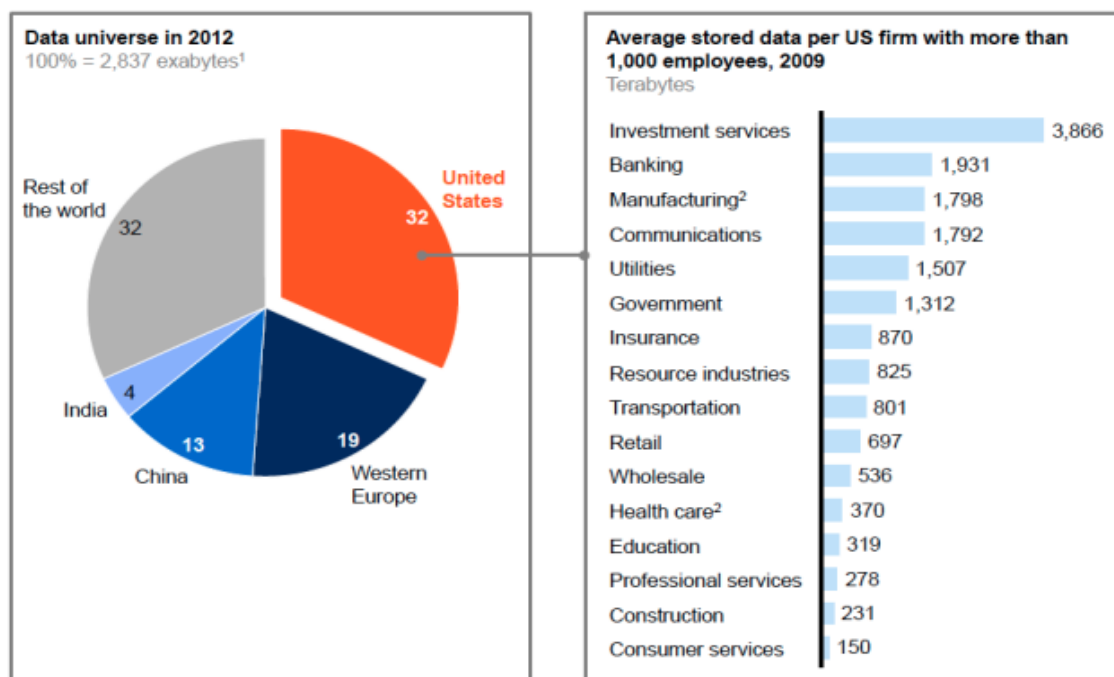
<sup>6</sup> <http://www.pewinternet.org/2012/02/24/privacy-management-on-social-media-sites/>

## 6. Prednosti i mane Velikih podataka

Veliko istraživanje instituta McKinsey Global iz 2011. donosi pregled važnosti Velikih podataka u svijetu na temelju prikupljenih podataka iz 2009. Ovim istraživanjem prikazuju kako su Veliki podatci u svim sektorima ekonomskog društva (vidi sliku 1.) pa tako njihovu zastupljenost pronalazimo u proizvodnji, vladama, komunikaciji i medijima, procesima tehničke izrade, bankovnom sektoru, zdravstvu, osiguravajućim i investicijskim društvima, profesionalnim uslužnim djelatnostima, maloprodaji, obrazovanju, osiguranju, prijevozu, veleprodaji, komunalnim službama, industriji sirovina, potrošačkim uslugama i rekreativnim aktivnostima i građevini.

### Exhibit E6

#### The United States has one-third of the world's data



<sup>1</sup> One exabyte = 1,024 terabytes, nearly 2.5 times as large as US Library of Congress web archive (as of May 2013).

<sup>2</sup> The large number of firms in the manufacturing and health-care sectors reduces the available storage per company.

SOURCE: IDC; US Bureau of Labor Statistics; US Library of Congress; McKinsey Global Institute analysis

Slika 1: Zastupljenost velikih podataka i omjer Velikih podataka u svijetu (preuzeto od Foxman, 2013)

Iako su ovi podatci za tržište Sjedinjenih Američkih Država, koje se smatra najvećim područjem za prikupljanje Velikih podataka s 32% cjelokupnih podataka svijeta (Slika 1), dok zapadna Europa ima tek 19%, i dalje nam pružaju sliku o važnosti Velikih podataka. Ja u svom istraživanju nisam uspjela pronaći objavljenu sustavnu statistiku o Velikim podacima u Republici Hrvatskoj. Velikim podacima o korisnicima u Hrvatskoj vjerojatno raspolažu tvrtke poput Hrvatskog Telekomu što potvrđuje i njihova nova usluga za poslovne korisnike *Heat Maps* koja poslodavcima nudi podatke o kretanju potrošača pored njihovih poslovnica i neke osnovne podatke o njima poput dobi, prebivališta i dohotka (Hrvatski Telekom, 2016). Cilj ove usluge je da poslodavci ostvare bolju dobit na temelju ciljanih grupa, pogodnih lokacija za nove poslovne ili radnog vremena koje bi bilo prilagođenije kupcima. Hrvatski Telekom pri tom se poziva na anonimnost danih informacija. I Hrvatski Zavod za zdravstveno osiguranje bi moglo ostvariti prikupljanje Velikih podataka s obzirom da su prije nekoliko godina uveli elektroničke recepte, digitalne snimke nalaza i digitalnu kartoteku.

Kako su Veliki podatci prisutni u svim područjima ekonomije, oni čine skoro sve ljude korisnicima Velikih podataka ili barem potencijalnim izvorima za prikupljanje Velikih podataka. Stavljajući na raspolaganje naše podatke kroz različite društvene mreže, surfanjem stranicama koje sadrže kolačiće (eng. *cookies*), koristeći kartice vjernosti (eng. *loyalty cards*) u trgovinama ili jednostavno koristeći pametni telefon kroz različite besplatne aplikacije, postajemo korisnici Velikih podataka. Pri tome ne upotrebljavamo samo njihova rješenja, već smo i izvor za prikupljanje novih podataka. Facebook kao društvena mreža sudjeluje u stvaranju podataka kroz 600 milijuna aktivnih korisnika koji provedu 9.3 bilijuna sati na stranici, YouTube tvrdi kako ima 490 milijuna korisnika koji svake minute stvaraju po jedan novi video sadržaj. Koliko ljudi obuhvaća ova revolucija potvrđuje i činjenica da je u 2010. više od 4 bilijuna ljudi ili 60% svjetske populacije posjedovalo mobilni telefon, a od toga njih 12% pametni telefon. Također, treba uzeti u obzir i da se korištenje pametnog telefona povećava 12% godišnje (Manyika, J. et al., 2011).

Važnost upotrebe Velikih podataka može se iščitati iz istraživanja o mogućnostima Velikih podataka, instituta McKinsey Global, koje predviđa kako bi javni

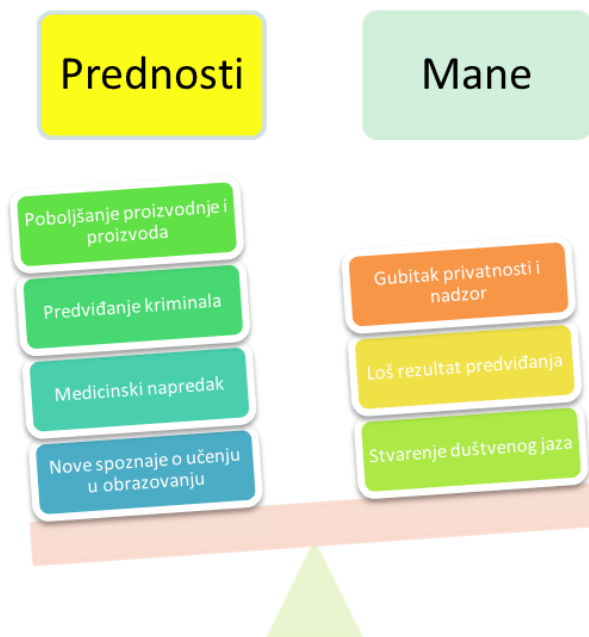
sektor Europske unije mogao ostvariti 250 bilijuna eura vrijednosti i postići rast od 0.5%, dok bi zdravstvo Sjedinjenih Američkih Država moglo ostvari 300 bilijuna dolara vrijednosti i postići rast od 0.7%, a industrijska proizvodnja bi mogla smanjiti troškove za 50% (Manyika, J. et al., 2011). Pomoću Velikih podataka prelazimo iz faze u kojoj su se stvari precizno promatrale te se odmičemo kako bismo vidjeli veću sliku iz šire perspektive. Tako na neki način, naš uzorak postaje sav intelekt (N= sav intelekt). Unatoč zamagljenoj slici i višeznačnosti, koju nismo tolerirali do pojave Velikih podataka, danas možemo pristati na taj kompromis jer nam on omogućava cjelovitiju sliku. (Mayer-Schönberger, Cukier; 2013).

Na sljedećim ću stranicama kroz primjere više autora pokušati pojasniti prednosti prikupljanja Velikih podataka u različitim sektorima. Tako, na temelju primjera autora Mayer- Schönberger i Cukier (2013), Needham (2013), Kocijan (2014), Cole, Nelson i McDaniel (2015), Silberman (2012), Hofstetter (2014) i Brücher (2013), Mayer-Schönberger i Cukier (2014), prednosti prikupljanja Velikih podataka možemo vidjeti prvo u rješavanju problema proizvodnje u poboljšanju same proizvodnje, potom u primjerima unutar medicine za predviđanja rizika od oboljenja i prilagođavanje liječenja svakom pojedincu, te na kraju u mogućnostima koje nam se otvaraju u obrazovanju kako bi smo bolje shvatili proces usvajanja znanja.

S druge strane se otvaraju mnoga pitanja za nas kao korisnike različitih usluga poput besplatnih aplikacija, elektroničke pošte, društvenih mreža čiji je sastavni dio i prikupljanje Velikih podataka. Treba uzeti u obzir kako ne treba slijepo vjerovati tehnologiji pa i u ovom slučaju razumjeti kako predviđanje sudbine nosi sa sobom rizik. Ako sve što nas okružje pokušamo predvidjeti, poput mogućih kriminalaca, nikad nećemo saznati jesmo li bili u pravu, upozoravaju Mayer-Schönberger i Cukier (2013). Isti autori te Hofstetter (2014) i Danielson (2012) govore i o mogućnosti nadziranja ljudi kroz tehnologiju pri čemu sami korisnici tehnologije gube kontrolu nad osobnim podacima i ne znaju što se s njima događa. Kao posljedica nadzora ograničena je sloboda pojedinca. Ni tvrtke nisu izuzete kada govorimo o manama prikupljanja Velikih podataka. Tako se prema istraživanju instituta McKinsey Global (2011) susrećemo s problemima nabave skupe tehnologije i nedostatkom radne snage tj. stručnjaka, ponajviše

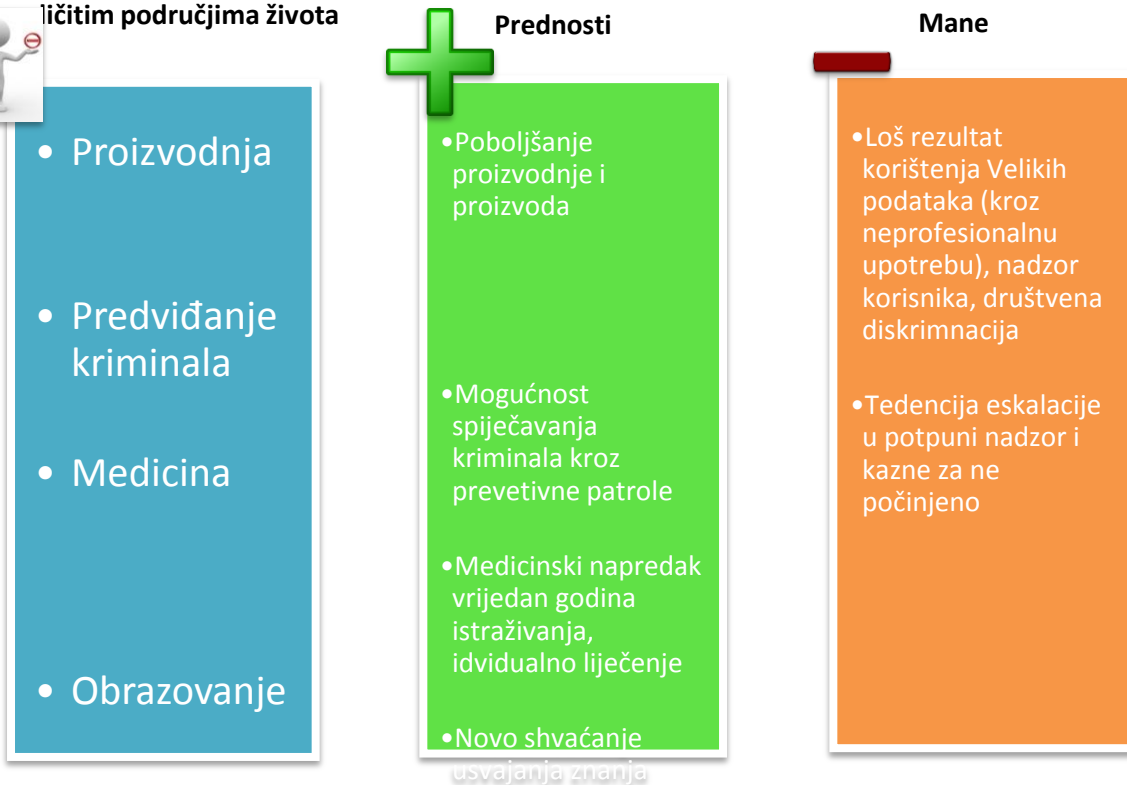


analitičara. Isti problemi pogađaju i javni sektor. Autorica Yvonne Hofstetter (2014) upozorava da ovi problemi ponekad rezultiraju lošim finalnim rezultatom koji je nastao prikupljanjem Velikih podataka najčešće u tvrtkama koje nisu spremne uložiti dovoljno novca pa kao rezultat prikupljanja Velikih podataka i njihove analize često dobiju nedovoljno kvalitetan završni produkt. Kako pozitivne primjene Velikih podataka mogu imati i svoje negativne strane, vidimo u grafu 2, u kojem su prikazane primjene Velikih podataka u različitim područjima života. Problem se može pojaviti i pri istraživanjima s Velikim podacima na što upozoravaju Boyd i Crawford (2012). Reno, Tene i Polonetsky (2012) smatraju kako su oni koji prikupljaju podatke u znatnoj prednosti nad onima kojima podatci pripadaju jer vlasnici podataka često nisu informirani što se s njihovim podacima dalje odvija, a što stvara dodatani jaz među onima koji imaju više sredstava i onima s manje društvene moći. Prema Lermanu (2013) ne treba niti zaboraviti one koji su u potpunosti izostavljeni iz revolucije. To su ljudi koji nisu korisnici sadržaja koji bi ih učinili izvorima Velikih podataka te na taj način bivaju izuzeti iz poboljšanja i napredaka koje Veliki podatci donose. Podjela prednosti i mana Velikih podataka može se još jednom iščitati iz grafa 1, a objašnjenje ovog prikaza bit će izneseno u potpoglavljima prednosti i mane.



**Graf 2: Prikaz prikupljanja Velikih podataka prikupljanja Velikih podataka**

## Prikupljanje Velikih podataka u različitim područjima života



Graf 2: Prikaz prikupljanja Velikih podataka u različitim područjima života

## 6.1. Prednosti

Kada autor Jeffrey Needham u podnaslovu svoje knjige (Needham; 2013) kaže: *Big Data! Big Bang!*, jasan nam je značaj Velikih podataka i njihovih beskrajnih mogućnosti. Danas smo svjesni kako Veliki podatci nisu samo rezervirani za društvene mreže i strojno generirane internetske stranice. Agencije i tvrtke pomoću njih pronalaze odgovore na pitanja koja si prethodno nisu mogli ni priuštiti, a Veliki podatci pomažu pri identifikaciji novih pitanja. U ovom ću poglavlju podijeliti prednosti u osnovna četiri potpoglavlja koja obuhvaćaju: napredak proizvodnje i poboljšanje proizvoda, medicinski napredak, predviđanje zločina i poboljšanje procesa obrazovanja na temelju Velikih podataka.

### 6.1.1 Poboljšanje proizvodnje i proizvoda

Proizvodnja je kao sektor od početaka računalne ere bila korisnik novih tehnologija i njihovih mogućnosti u širenju palete proizvoda. U devedesetima je tako doživjela veliki porast produktivnosti optimizacijom proizvodnje kroz uvođenje novih računala, a taj način i danas otvara mogućnost za nove napretke. Koristeći Velike podatke tvrtke koje se bave proizvodnjom su u mogućnosti ostvariti proizvode koji su razvijeni individualnije kako bi se približili korisniku, a isto u mogućnosti su i da pronađu nova rješenja u procesu proizvodnje kako bi on bio brži, efikasniji i jeftiniji. Pri tome tvrtke u ovom sektoru već raspolažu velikom količinom neobrađenih podataka, približno 2 eksabajta su bila spremljena prema informacijama iz 2010. godine. Podatci dolaze iz različitih izvora: iz same proizvodnje, iz sistema koji prate radni učinak proizvoda koji je već prodan (poput zrakoplovnih sistema) te se na taj način omogućava mikropraćenje proizvoda i njihovo poboljšanje (Manyika, J. et al., 2011).

Tako se povezivanjem prikupljenih Velikih podataka otvara novi svijet poboljšanja proizvodnje kroz Velike podatke, a uz pomoć prediktivne analitike, koja omogućava predviđanje događaja prije nego se oni dogode. Primjer tome je korištenje tehnologije kako bi se spriječile velike mehaničke ili strukturne pogreške. Na strojeve, motore ili infrastrukturu poput prenosnika stavljaju se senzori koji nagledaju obrasce koji se odašilju, poput topline, vibracije, pritiska i zvuka, a kako bi detektirali promjene

koje mogu signalizirati pojavu eventualnog problema. Do kvara u većini slučajeva ne dolazi odjednom već je on rezultat dužeg procesa. Oboružani podacima kao detektorima, korelacijska analiza i slične metode mogu identificirati specifične obrasce i izdajnički znak koji tipično iskrsne prije nego se nešto pokvari poput neobičnog zvuka motora, prekomjerne topline motora ili nečeg sličnog. Dakle, potrebno je samo tražiti obrazac kada nešto nije u redu. Primjećivanje abnormalnosti u ranoj fazi omogućuje sustavu da pošalje upozorenje kako bi se mogao instalirati novi dio ili kako bi se problem mogao riješiti prije nego se kvar zapravo dogodi. Cilj je identifikacija i time predviđanje budućih događaja. Međutim, ovakav način predviđanja neće pokazati je li do pregrijavanja došlo zbog izlisanog remena na ventilatoru ili loše pritegnute kape. Povezanosti pokazuju što, ali ne i zbog čega. Nasuprot Velikim podacima u ljudskoj je prirodi da traži veze i zapitkuje se zbog čega se nešto događa. Veliki podatci nam odgovor na to pitanje ne mogu ponuditi. Prihvaćanjem takvog načina razmišljanja iskoristit ćemo potencijal Velikih podataka za dobrobit proizvodnje (Mayer-Schönberger, Cukier; 2013).

Po prvi put zahvaljujući Velikim podacima proizvođači automobila si mogu priuštiti promatranje svog inventara rasprostranjenog po cijelom svijetu te prikupljanje petabajta podataka koji dolaze od senzora koji su danas ugrađeni u skoro svakom vozilu. Prije je ova akcija zahtijevala skupe centre za podatke koji su bili transferirani na superračunala, a što je bilo i vremenski zahtjevno. Sada pomoću Hadoop cluster<sup>7</sup> rasprostranjenog po svim potrebnim podacima troškovi su znatno manji. Uz to ovaj program omogućuje da se praktički preko noći dođe do jednostavnog rješenja zadanog problema (Needham, 2013).

Još jedna od brojnih upotreba Velikih podataka je korištenja podataka kako bi se odredilo gdje postaviti stanice za punjenje elektroničkih automobila. Za ovaj projekt je nužno da vozači mogu brzo i efektivno napuniti baterije, dok je za distributere struje bitno da se naponska mreža pri tome ne optereti. Svi podatci koji su se za ovaj projekt koristili imali su primarnu svrhu za neku drugu namjeru. To su bili podatci o vremenu, o stanju baterije u autu, o dobu dana, o stanju naponske mreže i sl. Spajanjem svih tih podataka stručnjaci su bili u mogućnosti odrediti mjesta na kojima je potrebno sagraditi

<sup>7</sup> Apache Hadoop je framework koji donosi distribuirano procesuiranje velikih blokova podataka koji nalaze na računalnim klasterima pomoću jednostavnih programskih modela.

još postaja za punjenje elektroničkih automobila. Time se pokazuje koliko su vrijedni Veliki podatci i kako je njihova primarna svrha za koju su bili prikupljeni tek dio onoga što nam još mogu pružiti kada ih odlučimo ponovno koristiti (Mayer-Schönberger, Cukier, 2013).

Brücher (2013) tako u svojoj knjizi i konkretno obrađuje primjer danskog proizvođača elektrana na vjetar koji analizu provodi pomoću Velikih podataka od samog početka analize do kraja, pri tome treba naglasiti kako proizvođač mora obratiti pozornost na 160 parametara. Upotrebom Hadoopa i MapReduce algoritma<sup>8</sup> omogućeno je da se izvođenja istih izračuna, koji su se stalno ponavljali, paralelno obrađuju na računalu. Konkretni pozitivni učinci velikih podataka pri ovom projektu bili su:

- na pitanja kupaca se moglo odgovoriti nakon nekoliko sati umjesto, kao prije, nakon 3 tjedna;
- za svaku pojedinu vjetrenjaču mogla se izračunati najbolja moguća pozicija, što je troškove po kilowatsatu proizvedene struje smanjilo;
- za svaku pojedinačnu vjetrenjaču moguće je precizno izračunati unaprijed dobit za godinu i troškove po kilowatsatu proizvedene struje, što je povećalo preciznost dobitka i smanjilo investicijske odluke voditelja pogona.

### 6.1.2. Predviđanje kriminala

Kroz širok spektar tehnoloških mogućnosti, a pomoću prediktivne analitike, prikupljanje velikih podataka omogućuje predviđanje zločina na temelju najrizičnijih predgrađa, ljudi i grupa. Ovlaštena strana tada je u mogućnosti temeljito proučiti odabrane podatke pomoću algoritama za predviđanje mogućnosti za počinjenje zločina (Mayer-Schönberger, Cukier; 2013).

---

<sup>8</sup> Kod MapReduce algoritama podatci se obrađuju u tri faze( *Map*, *Shuffel*, *Reduce*). Kroz ove faze izračuni nastaju paralelo i dijele se na više računala, kako bi se ti dijelovi obradili u zasebnim čvorovima, koji komuniciraju s drugim čvorovima i na kraju daju rezultate. Uspješnim ga čini to što se na ovaj način lako u *cluster*- ima može primijetiti greška.

Svoje početke, predviđanje zločina nalazi u sustavu LAPD, sustavu za predviđanje manjeg potresa, koji je rezultat više većih prijašnjih potresa. Zahvaljujući sličnim parametrima, u program su usadili 13 milijuna zločina iz zadnjih 80 godina. Time se moglo prodrijeti u prirodu zločina jer kada se negdje dogodi zločin, veća je mogućnost da će se u blizini ponovno dogoditi zločin, baš kao kod manjih potresa. Tim programom su se tek mogla odrediti žarišna područja zločina na temelju dotadašnjih podataka. Zbog toga su Sveučilište California i tvrtka PredPol poboljšali softver i algoritam kako bi predviđao moguća nova mjesta zločina za određen dan. Iako ovakva upotreba tehnologije nije još u svom punom jeku, neki počeci se nadziru. Tako policija grada Los Angeles koristi sustav predviđanja kriminala, a rezultate koje dobiju prediktivnom analitikom, koriste kako bi na ta područja koje je program smatrao izuzetno rizičnima slali patrole. Rezultat je 33% manje provala, 21% manje nasilnih zločina i 12% niža stopa otuđenja imovine u odnosu na prethodnu godinu. Pri tome treba naglasiti kako se ne radi o softveru koji se bazira na pojedincu i njegovoj zločinačkoj namjeri, već se radi o predviđanju rizičnih područja koja se posebno nadziru. Pomoću Velikih podataka, policija grada Durham u Engleskoj, na sličan se način bori protiv suzbijanja prevara osiguranja (Rijmenam, 2014).

### 6.1.3 Medicinski napredak

Kada se podatci višenamjenski koriste, onda njihova primarna svrha nije jedinstvena već se koriste i za daljnje analize. Tako su se pomoću programa praćenja novčanice Hanka Eskina prikupljeni podatci kasnije iskoristili kako bi se pratilo širenje gripe (Kocijan, 2014). Ovaj model praćenja novca nalazimo i u Hrvatskoj pod imenom *Gdje je moja kuna*. Iako je projekt u svojim početcima, jer zahtjeva određenu financijsku i marketinšku potporu, moguće je da jednog dana bude upotrebljen na sličan način (Požega, 2014).

Primjerice u medicinskom području istraživanja, znanstvenici i medicinski stručnjaci su u mogućnosti promatrati više pacijenata pri većoj brzini i točnosti u nastojanju da povećaju mogućnosti istraživanja. Jedan od takvih primjera je proučavanje globalnih infekcijskih bolesti. Puno je lakše spriječiti neku zaraznu bolest ako znamo na

kojem je geografskom području izbila. Do sada je samo dva posto zaraznih bolesti bilo geografski određeno. Razlog tome leži u tome što su znanstvenici prije određivali geografsko područje na temelju područja gdje su promatrali bolest. Veliki podatci omogućuju ubrzanje ovog procesa time što se podatci aktualiziraju u realnom vremenu i to svaki put kada je prijavljeno novo pojavljivanje bolesti. Ovi podatci se koriste zatim kako bi se razvila mapa predviđenih rizika. Cole, Nelson i McDaniel (2015) tako smatraju kako bi se u budućnosti mogla razviti cijela baza podataka o okolnostima bolesti.

Spajanjem različitih skupova podataka još su 2011. godine istraživači danskog Instituta za rak na temelju prethodno prikupljenih podataka o svim mobilnim pretplatnicima u Danskoj, podataka o dohotku i stupnju obrazovanja te na temelju podataka o oboljelima od raka centralnog živčanog sustava pokušali naći odgovor povećava li mobilni uređaj mogućnost dobivanja raka. Rezultat studije nije pokazao veliku vezu između raka i vlasnika mobilnih uređaja (Kocijan, 2014). Veliki podatci se pokazuju kao velika vrijednost kod istraživanja raka u današnjem svijetu omogućavajući spajanje područja prikupljanja podataka. Godinama je ovo istraživanje bilo limitirano na istraživanju na životinjama ili promatranju dijagnosticiranih pacijenata, a jedna i druga metoda su bile jako spore. Kao rješenje postavio se napravljeni algoritam koji očitava vrijednosti iz laboratorijskih istraživanja, koristi programe za čitanje dokumenta i ekstrahira značajne podatke te pretražuje različite baze podataka kako bi stekao informacije za pacijente dijagnosticirane sa sličnim stanjima. Zahvaljujući upotrebi podataka na ovaj način stručnjaci su u mogućnosti pronaći pacijente koji imaju sličan genetički kod, dijagnozu, laboratorijski izvještaj i druge različite sličnosti. Isto tako su u mogućnosti saznati koje su metode bile uspješne pri liječenju, a koje nisu. Pri tome novom pacijentu pristupaju s mnogo manje višestrukih mogućnosti rješenja pri liječenju i koriste konkretna rješenja koja su sužena prema točnosti. Spajanjem dosadnih i glomaznih podataka veliki podatci postaju najnapredniji alat u području medicine (Cole, Nelson, McDaniel; 2015).

Vrliki podatci mogu biti od pomoći i pojedincima s kroničnim bolestima. Tako postoji projekt unutar tima istraživača UbiComp koji se bavi mjerenjem funkcije pluća.

Ovo mjerenje se odvija pomoću pametnog telefona tako što se koristi mikrofoni ovog uređaja kako bi se izmjerila „buka“ koju korisnik stvara i na taj način pratilo kronično stanje bolesnika što omogućuje najdetaljnije i najefektivnije praćenje do sada. Nakon nekog vremena prikupljanja ovih podataka, podatke je moguće koristiti i za poboljšanje terapije ili detektiranje razloga zbog čega određena grupa bolesnika ima lošije posljedice (Silberman, 2012).

Veliki podatci omogućuju širok spektar pretraživanja baza podataka te potpomažu umjetnu inteligenciju, a dokaz je i priča o Watsonu, stroju umjetne inteligencije koji je razvijen u IBM kao računalo koje je sposobno odgovarati na pitanja prirodnim jezikom. Nakon razvojne faze, odlučili su ga testirati na kvizu *Jeopardy!*. Kao natjecatelj kviza Watson je raspolagao cijelom Wikipedijom i Enciklopedijom Britannicom. Ovaj se stroj pokazao kao jako uspješan u rješavanju problema i osvojio prvo mjesto i nagradu od milijun dolara. Njegova moguća daljnja upotreba može biti upravo medicinska analiza pri kojoj će davati brze odgovore na temelju svih pretraženih baza o liječenju i stanjima pacijenta ili kao umjetni manager za investicije koji je u mogućnosti svakodnevne kvalitativne, tekstualne tokove podataka u realnom vremenu interpretirati, kako bi donio odluku za najbolju investiciju (Hofstetter; 2014).

### 6.1.4. Obrazovanje

Proces obrazovanja može se usporediti s proizvodnom trakom: materijali za nastavu se mogu mijenjati i nastavnički procesi se mogu prilagođavati, ali usprkos svoj inovativnosti i predanim učiteljima, obrazovanje i edukacija se baziraju na prosječnim vrijednostima pri čemu se ne uzimaju u obzir individualne sklonosti, sposobnosti ili potrebe. Kada je usvajanje znanja u pitanju još smo u dobu industrijske revolucije kao da nas se ne tiče što sve ostalo uzima svoj tijek i kreće naprijed. Veliki podatci bi u ovom području mogli omogućiti bolju povratnu informaciju i individualan pristup usvajanju znanja. Na temelju vjerojatnosti koje se daju predvidjeti Velikim podacima, možemo ostvariti nove spoznaje o tome kako ljudi uče, te ih poboljšati (Mayer-Schönberger, Cukier; 2014).



Otvoreni online tečajevi, aplikacije za učenje i ankete za ispitivanje učenika o njihovom učenju, nude novi uvid u usvajanje znanja i obrazovanje. Iako se ovo većinom odnosi na učenje koje pojedinac sam potakne, ne znači da ne može doprinijeti spoznaje koje mogu pomoći samom školstvu. Nova tehnologija nam pomoću Velikih podataka pruža mogućnost da shvatimo koji dio lekcije je bio posebice težak, je li neki dio gradiva preskočen zbog ne razumijevanja, dosade ili umora, možemo shvatiti što je općenito teže određenoj skupini ljudi ili kako se učenje jezika razlikuje ovisno o tome koji je učeniku materinski jezik. Tako je na primjer tim iz Duolinga<sup>9</sup> otkrio kako ako nekoga, kome je materinski jezik španjolski, želimo naučiti njemački, ne treba koristiti pri početku posvojnu osobnu zamjenicu ovo koju govornik ne poznaje u svom jeziku pa mu ona stvara zbrku. Spoznajom da upotrebu te zamjenice treba uvesti tek nakon nekoliko tjedana, smanjili su broj onih koji su tečaj prekidali. Sve ovo prije nije bilo moguće jer je nastava najčešće bila frontalna i s povratnom informacijom koja je standardizirana za svih. Sada nam se otvaraju nova vrata u svijet učenja. I dalje su učitelji ti koji će na tom putu imati glavnu riječ, ali zaključci Velikih podataka mogu im pomoći kako bi poboljšali način usvajanja znanja (Mayer-Schönberger, Cukier; 2014).

### 6.2. Mane

Nakon što pomislimo kako nam prikupljanje velikih podataka donosi olakšanje u životu i svim primjerima pokazuje kako ćemo našu svakodnevicu učiniti boljom i naprednijem, Mayer-Schönberger i Cukier (2013) nas uvode u poglavlje rizika na nimalo idiličan način: povlačeći paralelu između Satsija u istočnoj Njemačkoj i mogućnosti kontrole ljudi kroz velike podatke.

Podatci koje omogućujemo poduzećima da koriste od nas ili koji se jednostavno prikupljaju našim digitalnim aktivnostima, ne moraju u svojoj suštini biti osobne informacije, ali mogućnosti koje nude veliki podatci, omogućuje procesuiranje podataka tako da se spoje s osobom čiji su. Pri svojoj usporedbi s istočnom Njemačkom autori naglašavaju kako je nadgledanje postalo lakše, jeftinije i moćnije. Mogućnost skupljanja

---

<sup>9</sup> Duolingo je portal za učenje stranih jezika, dostupan kao mobilna aplikacija i kao web platforma.

privatnih podataka je često ugrađena u alate koje koristimo svakodnevno, od internetskih stranica preko aplikacija. Iako je pri toj usporedbi jasno da nećemo zbog ovakvih podataka ići u zatvor i dalje nadgledanje prodire u sve aspekte našeg života. Pri tome ne mislimo samo na privatne tvrtke poput Google, već i vlade. Jedan od primjera je NSA u Americi, za koju se smatra da pretresa i zadržava 1,7 bilijuna elektroničke pošte, telefonskih razgovora i druge svakodnevne komunikacije, prema istraživanju Washington Posta iz 2010. godine (Mayer-Schönberger, Cukier; 2013). Unutar ovog poglavlja osvrnut ću se na to kako prikupljanje Velikih podataka može poći po zlu i na koje mane, prema mojim spoznajama, treba uzeti u obzir, pri čemu ću mane podijeliti na gubitak privatnosti i nadzor, stvaranje društvenih podjela te loš rezultat predviđanja.

### 6.2.1. Gubitak privatnosti i nadzor

Pitanje koje se postavlja pri tako velikom broju podataka je naravno činjenica da svi ne mogu biti obrađeni, pa zbog čega prikupljaju i dalje? Kada govorimo o nadgledanju, u tom je području došlo do promjene tehnika i načina razmišljanja. Dok je u svijetu malih podataka bilo bitno upoznati pojedinca kojeg se nagleda, danas se to odvija u maniri Googla i Facebooka, tako što se pojedinca gleda kroz njegov skup osobnih informacija, čime se spremanjem podataka omogućuje u potrebnom trenutku da se analiziraju podatci vezani za jednog individualca na temelju njegovih veza (Mayer-Schönberger, Cukier; 2013).

Većina kompanija, kojima su stručno područje Veliki podatci, koriste se frazama poput *bolja usluga našim kupcima, ciljani oglasi i oglasi koji su relevantniji našim korisnicima*. Sve se bazira na jednostavnom temelju poznavanja potrošača i na jednostavnom principu besplatne usluge za koju najčešće korisnici bivaju obasuti oglasima ili se informacije o korisnicima koriste za druge analize. Pri tome je većina korisnika zadovoljna ovakvim uvjetima. Naravno, nikad ne postoji garancija kako će se sakupljeni podatci koristiti isključivo u promidžbene svrhe jer se jednom prikupljeni podatci mogu koristiti na razne načine. Čak i sam način oglašavanja time postaje opasniji. Dok je nekad oglašavanje bilo jednosmjerno, vidjeli bismo oglas i ignorirali bismo ga,

danas, iako ga možemo ignorirati, ova funkcija se odvija u oba smjera, ignorirajući oglas i dalje ne znamo koje informacije oglašivač u tom trenutku prikuplja o nama. Zabrinjavajuće je da sve postaje osobno. Dok su se nekada reklame zasnivale na analizi šire populacije one se danas zasnivaju na osobnoj analizi pojedinca. (Reno, 2012)

Na primjeru koji je poslužio u dobre svrhe u prethodnom potpoglavlju, možemo vidjeti koliko smo ranjivi kao potrošači kada govorimo o Velikim podacima. Jedna osoba je s javnog računala u kafiću slala prijeteću elektroničku poštu s anonimnog korisničkog računa. Pri tome ga je policija mogla detektirati, bez pomoći pružatelja internetskih usluga, jednostavno na temelju analize oglasa koji su se prikazivali za vrijeme njegovog korištenja računala. Budući da je korisnik ostao prijavljen na stranici Facebook dok je slao poštu, nadležni su bili u mogućnosti pronaći njegov javni profil, što ih je na kraju dovelo do osumnjičenog (Danielson, 2012).

Kako je već pojašnjeno u poglavlju o prednostima Velikih podataka, u današnjem svijetu teži se k tomu da se proizvodi personaliziraju, kako bi nam život učinili lakšim i boljim. U smjeru te revolucije razvija se i Internet stvari, koji bi trebao obogatiti našu svakodnevicu. Međutim, treba biti oprezan kada se čini kako je sve podređeno nama. Tako na prvu zvuči smiješno da je upravo tvrtka Google kupila postavljača termostata Nest Labs ne bi li spojili Google i termostate. Ovaj sustav se bazira na algoritmima za sustave grijanja koji uče i prilagođavaju se. Karakteristike ovog algoritma mu omogućuju da nauči preference ukućana o klimi u sobama. Pri tome tehnologija radi sama za sebe bez manualnog namještanja vlasnika. I dok nekima to zvuči ugodno, onima upućenijima u Velike podatke postaje jasno da osobnije od toga ni ne ide. Cijela je kuća opremljena senzorima kako bi se na temelju vlažnosti zraka moglo zaključiti, jesu li ukućani kod kuće ili ako se možda odvijaju tjelesne aktivnosti kroz koje se još povećava vlažnost zraka. Nije teško zaključiti kako na temelju ovih informacija, onaj koji prikupi podatke, zna sve o našim navikama svakog dana (Hofstetter; 2014).

Tako danas govorimo o Internetu stvari koji ulazi u naše živote pod krilaticom da će optimizirati naš život. Imamo aute koji nam ne dopuštaju da odvežemo sigurnosni pojas kada se parkiramo u rikverc, bez obzira što ništa ne vidimo zavezani. Tu pripadaju i već spomenuti sustavi samoupravljačkog grijanja, hladnjaci koji se same pune i sl. Svi ti

uređaji pretvaraju naš svijet u nešto novo, dok je prije česta rečenica zastrašivanja bila: *Bog vidi sve*, danas se ona pretvorila u *NSA vidi sve*, *Google zna sve* i *Apple prisluškuje sve*. Pod krilaticom da je sve to kako bi se upravo građane zaštitilo te im se pružilo izvrsno temeljno pravo sigurnosti, postavlja se pitanje koliko je čovjek danas slobodan? Kada govorimo o tvrtkama Google, Apple ili Facebook nemamo taj prizvuk nadgledanja kao u nekim prijašnjim političkim režimima, kao da ih jednostavno smatramo poštenima. Ipak su oni ti koji su nas globalno povezali, a pri tom uništili pojam prijateljstva. Učinili su nas mobilnima i u isto vrijeme nam ukinuli anonimnost. Kada se spoje pametni telefoni direktno s kreditnim karticama, što se već počelo provoditi u djelo, tada će se naš trag bez premca svugdje pratiti. Interesantno je da nitko ne postavlja pitanje, kako bi ovi samoprovzani filantropi na čelu tvrtki jednostavno mogli promijeniti stav. Ili da bi država prema zakonskoj moći jednoga dana mogla zatražiti sve podatke. Ne radi se više samo o prediktivnoj analizi potrošačkih navika, radi se o samom životu. Hoće li nekome na temelju podataka biti omogućeno zdravstveno osiguranje, operacija, koja ima ili nema koristi, pa na kraju krajeva i samo rođenje i smrt. Kada stvari postavimo tako jasno u središte samog života nitko ne može reći kako ga se ovo ne tiče (Hofstetter; 2014).

U svijetu malih podataka pristajali smo na generalna predviđanja poput onih da se u automobilu vežemo sigurnosnim pojasom kako bi smanjili moguće ozljede pri nesreći, poput sigurnosne provjere pri ukrcanju u zrakoplov ili pak da muškarce koji su prešli pedesetu podsjećamo na nužnost provjere prostate jer spadaju u rizičnu grupu. Generalizacija kod analize podataka s Velikim podacima poprima nov individualan oblik pa tako primjerice polica osiguranja neće biti bazirana na temelju jedne generalne grupe, već će se istraživati povijest onog pojedinca kojem se prodaje određena polica osiguranja. Obećanje koje nam daju Veliki podatci je da profiliranje postaje bolje, manje diskriminirajuće i personalizirano. Dok je prihvatljivo takvu analizu obraditi kako bi zaustavili neželjene radnje, s druge strane je opasno ako koristimo ove prognoze kako bi odlučili da li je netko sposoban i da li ga je nužno kazniti za još nepočinjeno. Primjer predviđanja kriminala u poglavlju prednosti pokazuje svoje dobrobiti, no treba uzeti u obzir da se cilja k tome da je moguće u potpunosti predvidjeti individualnu kriminalnu namjeru za koju bi zatim slijedile i kazne bez da je kriminalna radnja počinjena. Na tragu

toga je FAST ( *Future Attribute Screening Technology*) program koji na temelju kretnji i gestikulacije predviđa je li individualac koji se promatra terorist (Mayer-Schönberger, Cukier; 2013).

Bitno je shvatiti kako analitičari neće reći kako su njihova predviđanja na temelju Velikih podataka savršena, već da buduće ponašanje ima određenu vjerojatnost. Koristeći takve predikcije te određujući za njih kazne ne damo sudbini da se odvije i pri tome nikad sigurno ne bismo mogli reći kako jesmo ili nismo bili u pravu (Mayer-Schönberger, Cukier; 2013).

*Možda s takvim sistemom bi društvo bilo sigurnije ili efektivnije, ali esencija onoga što nas čini ljudima - naša mogućnost da odaberemo naše postupke i da se držimo odgovornima za njih - bila bi uništena*<sup>10</sup> (Mayer-Schönberger, Cukier; 2013:162).

### 6.2.2. Loš rezultat predviđanja

Iako smo kod poduzeća vidjeli kako Veliki podatci mogu odvesti do prednosti, smanjiti troškove, poboljšati proizvodnju i rješavanje problema, ipak se i oni bore s nekim manama koje Veliki podatci donose sa sobom. Kao prvo imaju problem nedostatka tehnologije koja je primjerena Velikim podacima. Ovaj problem obuhvaća i javni sektor pri korištenju Velikih podataka. Novi podatci zahtijevaju drugačije tehnologije od onih koje su se prije koristile pogotovo kada se govori o poboljšanju proizvoda. Zatim postoji i potreba za tehnološkim kadrom, kojim se često neučinkovito upravlja, svrstavajući analitičare i tehničare u zasebne sektore pri čemu nisu uvršteni u donošenje odluka i imaju ovlasti tek u zadacima koji su im postavljeni kako bi ih točno odgovorili. Ovim načinom prikupljanje Velikih podataka odvija se na nepotpun način što može dovesti i do krivih rješenja. Naravno, kako i kod ostalih područja u kojima se prikupljaju Veliki podatci, niti ovo područje nije oslobođeno problema privatnosti. Stoga možemo zaključiti da ih izazov rješavanja problema korištenja podataka kupaca proizvoda i dalje čeka (McKinsey, 2011).

<sup>10</sup> U originalu: Perhaps with such a system society would be safer or more efficient, but an essential part of what makes us human- our ability to choose the actions we take and be held accountable for them- would be destroyed. Big data would have become a tool to collectivize human choice and abandon free will in our society.

Jedan primjer kako se jednostavno nepotpunom upotrebom Velikih podataka može utjecati na život korisnika je polica osiguranja za automobil, koje bi se moglo u budućnosti računati pomoću navigacijskih uređaja u automobilima kako bi se odredile navike vozača. Većini je poznato da navigacijski uređaji nisu najpouzdaniji, što bi i mnogi stručnjaci potvrdili. Pri tome, dođe li do ovakvog izračunavanja osiguranja, jednostavna greška koja se može dogoditi je da vas uređaj locira na krivom mjestu, da vozite na autocesti, a da je vaša lokacija u navigacijskom uređaju u mjestu do autoceste gdje je ograničenje 40 km/h. Na jednom se javlja problem da prema prikupljenim podacima vozite trostruko brže u mirnoj zoni. Tu se javlja problem nereda spominjan u poglavlju o prikupljanju Velikih podataka da pri mjerenju stroju koji analizira podatke nisu dostupni egzaktni podatci. To je istina današnjice. Kada ne pričamo o primjeni podataka u vojne ili znanstvene svrhe već u gospodarske ovakvi problemi se događaju. U gospodarstvu je prekratko vrijeme obrazovanja za ovakve probleme. Nedostaje vrijeme i za obradu probleme, a nedostaje i nužnost kako da se problem kvalitetno obradi. Kvaliteta je skupa. Možemo ovaj problem lako usporediti s proizvodnjom odjeće na bliskom istoku. Analiza podataka sama po sebi nije loša, ali za nju je potreban egzaktn znanstveni rad, jasni koncepti i matematički znalci koji su sposobni prodrijeti u dubini Veliki podataka (Hofstetter, 2014).

Rizik jednostavnog shvaćanja prikupljanja podataka bez prave znanstvene perspektive pojavljuje se i kod istraživanja. Veliki podatci se smatraju boljim kako smo to i do sada vidjeli jer pružaju veću perspektivu, uključujući više potrošača. Pitanje je koliko je taj pristup ispravan. Uzet ćemo za primjer Twitter koji se do sada koristio kako bi se istraživali različiti uzorci poput medijskih angažmana, političke popularnosti i sporazumnih interakcija. Teško možemo reći da je Twitter sva populacija te ne možemo izjednačavati korisnike Twittera s ljudima. Niti su kao takvi predstavnici globalne populacije, ima ljudi koji uopće nemaju Twitter, isto tako ima ljudi koji imaju više Twitter računa ili Twitter računa iza kojih stoji više ljudi. Nadalje i *tweet*-ovi koji su dostupni nisu nikada potpuni jer često nedostaju kada se preuzmu iz baze. Zato treba obratiti pozornost kako su *tweet*-ovi možda reprezentativni, ali ne nužno. (Boyd, Crawford; 2012)

### 6.2.3. Stvaranje društvenog jaza

Još jedan problem koji se pojavio s Velikim podacima je već odavno nejednaka razlika između individualca i velikih poduzeća. Već inferiorni individualci sada postaju još inferiorniji. Iako podatci nekog pojedinca jesu njegovi i o njemu, oni nisu u njegovu vlasništvu. I ne samo da nisu u njegovom vlasništvu, već nad njima nema nikakav pregled niti informaciju o tome što se s njima događa. Ovim se gubi granica između onog što mi kao potrošači ili konzumenti neke opcije radimo i što nam se prediktivnom analizom Velikih podataka nalaže da ćemo tek napraviti. Što se tiče ovog problema jasno pokazuje primjer koji je 2012. godine iznio New York Times o prediktivnoj analizi jednog dućana koji je na temelju različitih parametara donosio zaključke o tome je li neka žena trudna i na temelju toga joj slao akcijske letke kući. Priča govori o ocu koji je na ovaj način saznao da mu je maloljetna kćer trudna. To je granica koju želimo zadržati za sebe, a ne da odluku kada će naša obitelj tako nešto saznati, za nas donosi nekakvo poduzeće na temelju akcijske ponude (Tene, Polonetsky 2012).

Dok se svi autori iz svog gledišta fokusiraju na probleme koji bi mogli naštetiti nama kao korisnicima digitalnih pomagala, a time i sustvarateljima velikih podataka, Jonas Lerman (2013) se okreće jednom drugom gledištu obraćajući pozornost na one, koje ova revolucija u potpunosti zaobilazi, a time i diskriminira. Veliki podatci stvaraju rizik i onima koji nisu dio njih te čiji se podatci ne prikupljaju i ne analiziraju. Istina je da bilijuni ljudi ostaju van ove revolucije jer nisu dio aktivnosti koja uključuje Velike podatke i naprednu analizu kojom se ti podatci obuhvaćaju. Razlozi zbog čega su ti ljudi isključeni iz ove revolucije nisu važni. Može to biti siromaštvo, mjesto na kojem žive, stil života ili slično. Pri tome se autor ne referira na nekakav grad u Africi čije ime nitko ne zna, već sam primjer najsiromašnijeg grada u Americi koji je dovoljan da bi se pokazalo kako se lako ljudi isključe iz društva. Kada poduzeća nemaju na temelju čega prikupljati informacije o potrošnji u određenom području, moguće je da neće otvarati poslovnice, čime se ne ukida samo mogućnost izbora za potrošače, već i radna mjesta. Sve veće oslanjanje uprave i vlade na Velike podatke može dovesti do toga da zbog manjka

informacija bude shvaćeno kako javne ustanove nisu potrebne na tom području. Veliki podatci nisu samo prijetnja privatnost, već bi mogli ugroziti političku i socijalnu jednakost tjerajući ranjive u inferioran status (Lerman, 2013).



## 7. Zaključak

Naš se svijet iz dana u dan mijenja. Tehnološki napredak pritom nema više glavnu riječ kada spominjemo tehnologiju i digitalizaciju. S usponom Velikih podataka informacije mu stoje rame uz rame. One mijenjaju naš svijet na nov i inovativan način pri čemu je nužna naša prilagodba razmišljanja u novom smjeru. Datafikacija je usmjerila to razmišljanje u novom pravcu pretvarajući stare podatke u izvor novih mogućnosti i zaključaka.

Kako Veliki podatci mijenjaju naš svijet iz temelja, tako nas navode da preispitujemo što je privatnost i koliko je privatnosti ostalo u ovom novom svijetu podataka. Države polako svojim zakonskim regulativama pokazuju nužnost zaštite korisnika te nalaženje pravog puta kako iskoristiti podatke, ali istovremeno zadržati privatnost. Pri tome neki korisnici nisu ni svjesni što se događa s njihovim podacima i u koje su sve procese oni uključeni.

Ništa u životu ne dolazi samo kao dobro ili kao loše, a upravo tako je i s prikupljanjem Velikih podataka. Dok s jedne strane ono nudi mogućnost poboljšanja proizvodnje i samog proizvoda, sprječavanje kriminala na temelju predviđanja, veliki napredak u medicini uz mala ulaganja i nove načine kako shvatiti ljudsko usvajanje znanja. S druge strane javlja se mogućnost gubitka privatnosti kroz dostupnost podatka, nadzor od institucija i tvrtki, isto tako nedostatkom ili ne ulaganjem u pravu tehnologiju i stručne kadrove produkti zaključaka Velikih podataka mogu biti neprecizni i štetni, a inzistiranje na stalnoj upotrebi Velikih podataka u svim sektorima, može izolirati one koji nisu korisnici istih.

Iako sam pri početku pisanja ovog rada smatrala kako prikupljanje Velikih podataka dovodi do gubitka privatnosti i lak nadzor korisnika, nakon boljeg upoznavanja teme jasno mi je kolike nam se mogućnosti otvaraju za poboljšanje različitih sektora života, pri čemu se u nekim važnim područjima kao što su medicina i obrazovanje može napraviti napredak za koji su prije bile potrebne godine istraživanja. Veliki podatci danas

otvaraju nove svjetove mogućnosti pomoću svojih inovativnih rješenja. Na žalost niti oni nisu imuni na moguću zloupotrebu koja bi za posljedicu imala stalni nadzor i gubitak privatnosti. U budućnosti je na vladama da nađu bolju zakonsku regulativu kada govorimo o Velikim podacima, da je prilagode vremenu i tehnologiji.

U cilju poboljšanja situacije treba pronaći način kako da podatci postanu anonimni, a da se i dalje mogu koristiti za napredak. Isto tako potrebno je educirati građane i osvijestiti im što su to Veliki podatci i kakav utjecaj na život pojedinaca imaju. Na kraju bismo trebali razmotriti i činjenicu da je dio populacije izostavljen iz ove revolucije te pronaći način kako bi se i njih uključilo u proces napretka.

Možda bismo upravo na ova pitanja trebali potražiti odgovore u Velikim podacima, moguće da bi analiza rezultirala novim inovativnim rješenjem. Do tada nam preostaje čekati što će nam sve Veliki podatci donijeti, a sigurno je da ćemo o njima još puno čuti.

## 8. Literatura

1. Ackermann, A. (2015). „EU-Datenschutz-Grundverordnung: Das sind die Neuerungen“. Datenschutzbeauftragter-info, 16.Dezember. dostupno na: <https://www.datenschutzbeauftragter-info.de> (pristupljeno 22.6.2016).
2. Ackermann, A. (2015). „EU-Datenschutz-Grundverordnung: Das sind die Neuerungen“. *Datenschutzbeauftragter-info*, 16.Dezember. Dostupno na: <https://www.datenschutzbeauftragter-info.de> (pristupljeno 22.6.2016).
3. „Age of Analytics“, Northwestern Journal of Technology and Intellectual Property dostupno na: <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1037&context=sais2015> (pristupljeno 22.6.2016.).
4. Boyd, D., Crawford K. (2011) „Six Provocations for Big Data, in A Decade in Internet Time“, Symposium on the Dynamics of the Internet and Society, dostupno na: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1926431](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431) (pristupljeno 22.6.2016.).
5. Brüchner, C. (2013). *Rethink Big Data*. Hamburg: MITP Verlags GmbH & Co. KG.62-66
6. Cole D., Nelson J., McDaniel B. (2015) „Benefits and Risks of Big Data“, Association for Information Systems AIS Electronic Library (AISeL) SAIS 2015, dostupno na: <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1037&context=sais2015> (pristupljeno 22.6.2016.).
7. Foxman, S.(2013). The US is home to one third of the world’s data—here’s who’s storing it. Quartz. dostupno na: <http://qz.com/104868/the-us-is-home-to-one-third-of-the-worlds-data-heres-whos-storing-it/> (pristupljeno 22.6.2016).
8. Hofstetter, Y. (2014). *Sie wissen alles*. München: C. Bertelsmann

9. Hrvatski Telekom. (2016) „Heat Maps“ dostupno na:  
<https://www.hrvatskitelekom.hr/poslovnict/medijske-digitalne-usluge/heat-maps>  
(pristupljeno 22.6.2016.).
10. Kitchin, R. (2014). *The Data Revolution: Big Data, open Data, Data Infrastructures & Their consequences*. London: SAGE Publications Ltd.
11. Kocijan, K. (2014). „Big Data: kako smo došli do Velikih podataka i kamo nas oni vode“. U: *Komunikacijski obrasci i informacijska znanost*. Zagreb: Zavod za informacijske studije, 37-62.
12. Kuhn, J. (2016). „Google darf Millionen Bücher scannen - endgültig“. *Süddeutsche Zeitung*, 18. April. dostupno na: <http://www.sueddeutsche.de/digital/zehnjaehriger-rechtsstreit-google-darf-millionen-buecher-scannen-endgueltig-1.2955967>  
(pristupljeno 22.6.2016.).
13. Laboratorij za sustave i signale. Fakulteta elektrotehnike i računarstva, Sveučilišta u Zagrebu. (2010) „Sigurnost na Internetu“ 17-19 dostupno na:  
<http://security.lss.hr/Novi-dokumenti/privatnost-na-internetu.html> (pristupljeno 22.6.2016.).
14. Lerman, J. (2013), „Big Data and Its Exclusions“ u 66 *Stanford Law Review* Online 55, dostupno na: <http://ssrn.com/abstract=2293765> (pristupljeno 22.6.2016.).
15. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A. McKinsey Global Institute. (2011) „Big data: The next frontier for innovation, competition, and productivity“, 15-27, 37-76 dostupno na:  
[http://www.mckinsey.com/~media/McKinsey/Business%20Functions/Business%20Technology/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI\\_big\\_data\\_full\\_report.ashx](http://www.mckinsey.com/~media/McKinsey/Business%20Functions/Business%20Technology/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx) (pristupljeno 22.6.2016.).
16. Mayer-Schönberger, V.; Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.
17. Mayer-Schönberger, V.; Cukier, K. (2014). *Lernen mit Big Data: Die Zukunft der Bildung*. München: Redline Verlag.

18. Narodne novine (2010) Ustav Republike Hrvatske. Zagreb: Narodne novine d.d.  
dostupno na: [http://narodne-novine.nn.hr/clanci/sluzbeni/2010\\_07\\_85\\_2422.html](http://narodne-novine.nn.hr/clanci/sluzbeni/2010_07_85_2422.html)  
(pristupljeno 22.6.2016).
19. Narodne novine (2012) Zakon o zaštiti osobnih podataka. Zagreb: Narodne novine  
d.d. dostupno na: [http://narodne-novine.nn.hr/clanci/sluzbeni/2012\\_09\\_106\\_2300.html](http://narodne-novine.nn.hr/clanci/sluzbeni/2012_09_106_2300.html) (pristupljeno 22.6.2016).
20. Needham, J. (2013). Disruptive possibilities: How Big Data Changes Everything.  
United States: O'Reilly Media
21. Požega, M. (2015), „Primjer informacijskog sustava u web 2.0 okruženju (Projekt  
praćenja kretanja novčanica hrvatske kune u Republici Hrvatskoj).Filozofski  
Fakultet,Zagreb 2015. dostupno na: <http://darhiv.ffzg.unizg.hr/6220/> (pristupljeno  
3.9.2016)
22. Reno,J.. (2012). „Big Data, Little Privacy“, CA Technology Exchange Insights  
from CA Technologies: Big Data. 24-31 dostupno na:  
<http://www.ca.com/us/~media/files/about%20us/catx/catx-big-data-oct2012.aspx>  
(pristupljeno 22.6.2016).
23. Rijmenam, M. (2014) „The Los Angeles Police Department Is Predicting and  
Fighting Crime With Big Data“ *Datafloq*, 14.April. dostupno na:  
<https://datafloq.com/read/los-angeles-police-department-predicts-fights-crim/279>  
(pristupljeno 22.6.2016).
24. Silberman,G. (2012). „Some Noise About Big Data“, CA Technology Exchange  
Insights from CA Technologies: Big Data. 5-8 dostupno na:  
<http://www.ca.com/us/~media/files/about%20us/catx/catx-big-data-oct2012.aspx>  
(pristupljeno 22.6.2016).
25. Tene O., Polonetsky J. (2013), „Big Data for All: Privacy and User Control in the  
Age of Analytics. dostupno na:  
<http://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1191&context=njtip> (pristupljeno 1.7.2016)

26. WPressUTexas (2015) University of Texas at Austin, dostupno na:

[http://wpressutexas.net/cs378h/index.php?title=Is "Notice and Consent" a viable continuing framework%3F](http://wpressutexas.net/cs378h/index.php?title=Is%20%22Notice%20and%20Consent%22%20a%20viable_continuing_framework%3F) (pristupljeno 1.7.2016)